Novel Technique for Developing Linearized Convex System Models from Experimentally Derived Data

Preprint

Aadil Latif, Richard Bryce, and Adarsh Nagarajan

National Renewable Energy Laboratory

Presented at the IEEE Power and Energy Society (PES) General Meeting
Atlanta, Georgia
August 4-8, 2019

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-72754
October 2019

Novel Technique for Developing Linearized Convex System Models from Experimentally Derived Data

Preprint

Aadil Latif, Richard Bryce, and Adarsh Nagarajan

National Renewable Energy Laboratory

Suggested Citation
Latif, Aadil, Richard Bryce, and Adarsh Nagarajan. 2019. *Novel Technique for Developing Linearized Convex System Models from Experimentally Derived Data: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5D00-72754. https://www.nrel.gov/docs/fy20osti/72754.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-72754
October 2019

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Novel Technique for Developing Linearized Convex System Models from Experimentally Derived Data

Aadil Latif, Richard Bryce, Adarsh Nagarajan
National Renewable Energy Laboratory
Golden, Colorado, United States of America

***Abstract*—This paper presents a novel technique for generating a convex system model from an experimentally derived data set which features variance among repeated measurements. The convex system model developed as a test case characterizes the dynamic system losses of a vanadium redox flow battery as a function of the active power output and the battery state of charge. The technique hinges on a pre-cleaning via clustering procedure which precedes the formation of a planar convex hull comprised of triangular simplices. The clustering procedure efficiently reduces the experimental data set while mitigating variance among repeated measurements and removes outliers. Ultimately, the lower evelope of the planar convex hull serves as the desired convex system model. The proposed technique reduces systematic model error which is otherwise present when directly developing a planar convex hull model based on an unreduced data set.**

***Index Terms*—clustering, system model, convex hull, piece-wise linear models, DB-scan, vanadium redox flow battery**

I. INTRODUCTION

Modeling systems which feature convex interdependence between variables is commonplace in power systems engineering. Further, linearization of the convex space enables tractable solutions to optimization problems and is a common extension of the premise [1]. Such modeling efforts, when based on experimental data which is naturally riddled with measurement error and variance within repeated measurements, can be a tedious and error prone. This is particularly true when the underlying relationship between the variables to be modelled is not known a priori, this fitting a metamodel via a least-squares method is not a perspicuous endevour [2]. Moreover, generating a set of linearized constraint equations which reasonably represents such a metamodel introduces further model error. Thus, a modeling technique for reducing the impact of the measurement noise while simultaneously yielding a set of linear constraints bounding a convex space is desirable.

Although a number of approaches exist in literature for development of nonlinear models that capture system dynamics with various degress of fidelity, the implementation of these models within optimization frameworks may not be tractable or even possible in some cases [3].

Most system dynamic models in reality are nonlinear, but not necessariliy convex . In practical implementations many of these non-convex functions are approximated as convex functions with reasonable accuracy [4], [5]. Economic dispatch of generation units in power systems domain is one such example. Cost function of generation units has typically been approximated as a quadratic function for optimization problems. In reality however, steam valve operation in thermal turbines makes the cost function highly non-linear and non-convex. Quadratic approximation makes the problem tractable for optimization solvers [6].

Piece-wise linear models have a number theoretically unfavourable properties such as discontinuity and undifferentiability, however, they are directly applicable in mix integer linear programming (MILP) problems which makes them useful in practical applications [7]. Herein, such a technique is developed and validated against a set of experimental data of the dynamic losses of a vanadium redox flow battery (VRFB) as a function of the active power output and the state of charge (SOC) of the system. The intent is to develop a storage system loss model that can in the future be embedded into ReOpt; A MILP framework based techno-economic analysis tool developed and maintained by National Renewable Energy Lab [8].

II. DATA AND METHODS

The data set used in this work has been sourced from cycle testing a utility scale redox flow battery system as it operates within a dsitribution network. In this work, all recorded measurements have been scaled by respective peak values to protect manufacturer's proprietary information. The storage system was charged and discharged at various levels of state of charge and active power over a period of three days, such that most of the state space was sampled. In this study various components of system losses have been summed and are referred to as 'total system losses' here after. Although in this paper storage system loss model development is the chosen use case, the proposed model development method is extendable to any data set that is predominantly convex. Figure 1 shows surface plot of scaled total system losses measured at various levels of active power and state-of-charge.

A. Formation of the convex hull

The convex hull C is formed by enveloping a given set G containing all measurement points, $p_i \equiv (P_m, S_m, Lm)$ using Equation 1. P_m, S_m, Lm are measured active power, state of charge and total system losses for the flow battery respectively. Next lower envelope of the convex hull is identified and used to define a piece-wise linear model for total system loss estimation. Algorithm 1 presents pseudo code algorithm developed and used for this purpose.

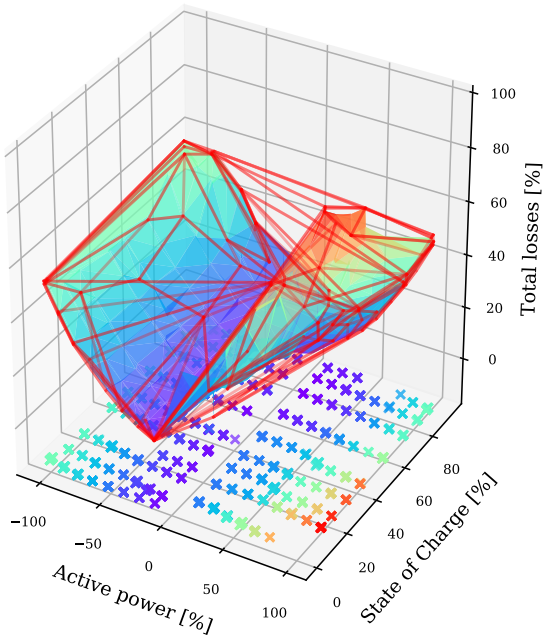

Fig. 1: Surface plot for total system losses measured for battery operated at various levels of active power and state-of-charge. Projection on the Z axis shows regions sampled during experimental run.

$$C \equiv H(G) = \left\{ \sum_{i=1}^{|G|} \alpha_i p_i \middle| (\forall i : \alpha_i \geq 0) \wedge \sum_{i=1}^{|G|} \alpha_i = 1 \right\} \quad (1)$$

For noisy experimental data sets convex hull forms on the extremities and may be heavily distorted by outliers. Thus

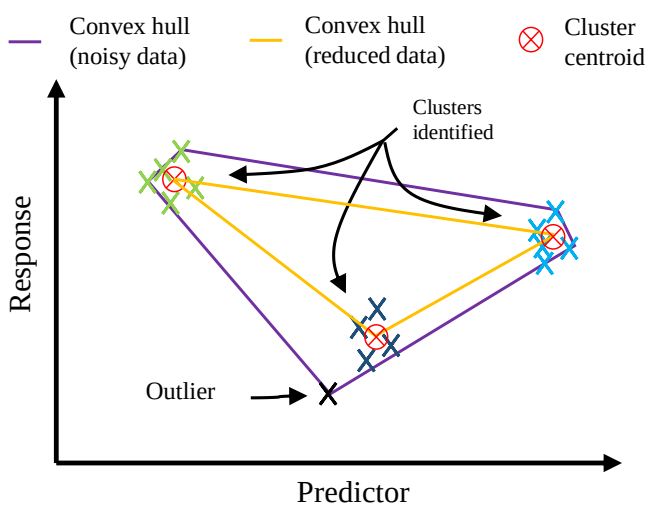

Fig. 2: The illustration shows why outlier and natural variance in measured data systematic under estimation if not cleaned.

using the formulated convex hull model would consequently result in systematic under estimation of system losses. In this work, a clustering based method has been proposed to mitigate errors associated with data noise and outliers. Figure 2 illustrates this problem using simple diagram.

Algorithm 1: Pseudo code for calculation of system losses from the convex hull model

Input: C, P_m, S_m
Output: L_e `// Estimated system losses`

p_{Dmin} = empty list;
$Losses$ = empty list;
$Dmin = \text{inf}$;
`// Identify the planes that contain` P_m, S_m
for ***each*** *hyperplane* $h_{abc} = a.P + b.S + c.L + d = 0$ ***forming*** *convex hull* C **do**
  Identify triplet (p_a, p_b, p_c) that bound the hyperplane h_{abc} ;
  Calculate K; projection of hyperplane h_{abc} onto $L = 0$ plane bounded by triplet (p_a, p_B, p_c) ;
  if *(P_m, S_m) lies within the region K* **then**
    Calculate losses l using the hyperplane h_{abc};
    Append l to $Losses$
if *length of list $'Losses' > 0$* **then**
  $L_e = \min(Losses)$ `// Return the loss value corresponding to the lower envelope of the hull`
else
  `// If` P_m, S_m `lie outside the hull` C
  for ***each*** *hyperplane h_{abc}* ***forming*** *hull C* **do**
    Identify triplet (p_a, p_b, p_c) that bound the hyperplane h_{abc} ;
    if *p_{Dmin}* ***in*** *triplet that bounds h_{abc}* **then**
      `// Identify point closest to measured values` P_m `and` S_m
      for ***each*** *point p_x* ***in*** *triplet* **do**
        Calculate distance D between $[P_m, S_m]$ and the project of point p_x on $L = 0$ axis;
        if $D < Dmin$ **then**
          $Dmin = D$;
          $p_{Dmin} = p_x$
  `// Extrapolate to estimate losses` L_e `using hyperplanes intersecting at the closest point` p_{Dmin}
  for ***each*** *hyperplane h_{abc}* ***forming*** *hull C* **do**
    if *p_{Dmin}* ***in*** *triplet that bounds h_{abc}* **then**
      Calculate losses l using the hyperplane h_{abc};
      Append l to $Losses$
  remove outliers from $Losses$;
  $L_e = \text{avg}(Losses)$

B. Clustering for outlier detection and data reduction

DBSCAN is a density based clustering algorithm which unlike k-means, does not require number of clusters as an input. It however, requires tuning of two parameters ϵ and n_{min}. Parameter ϵ defines maximum allowable distance between two points to be considered neighbors. For a neighbourhood to be considered a cluster, it should contain at least n_{min} points [9], [10]. When working with noisy data sets, such as the case here, one big advantage of using DBSCAN clustering algorithm is that it is able to detect outliers and disregard them when forming clusters.

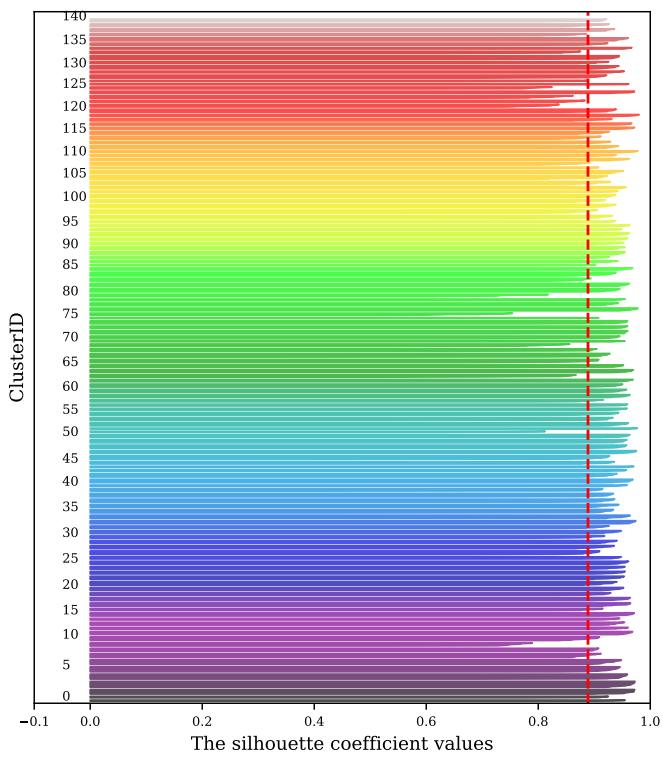

Fig. 3: silhouette plot for clusters formed using DBSCAN algorithm

Number of clusters (k) formed is dependent on values chosen for input parameters ϵ and n_{min}. Although no exact method exists to identify optimal values for the input parameters, silhouette plot is one method that is commonly used to fine tune parameters ϵ and n_{min} [11]. Silhouette score s_i is an indicator of how far a point is from a boundary between two clusters and ranges for -1 to 1 and is calculated using Equation 2.

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \qquad i \in (1, 2, ..., k) \tag{2}$$

Where a_i is the average distance of any data point i and all other data points within the same cluster and b_i is the average distance of any data point i and all data points with the cluster closest to the cluster than contains i. Values close to 1 indicate that the point if far from the decision boundary, while, values close to -1 indicate that a point may not belong to the right cluster.

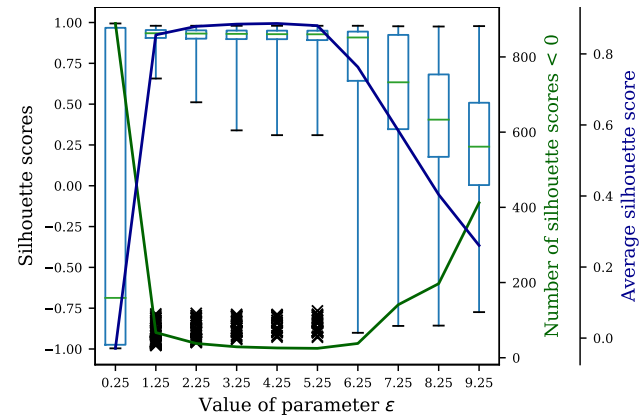

Fig. 4: The plot shows the impact of varying the the value parameter ϵ while fixing the value of parameter n_{min} to 3 on the silhouette scores.

In this work, a parametric study was performed and with value of ϵ varying from 0.25% to 10% and n_{min} incrementally changing from 1 to 5 points. A number of parameters average such as silhouette score, number of scores less than zero and variance winth the score etc. were used to identify values of parameters ϵ and n_{min} (Figure 4). In this study the values chosen as 4.25 and 3 respectively. Figure 3 shows the silhouette plot for these values. The number of clusters identified within the data set is 141 and the average silhouette score is 0.89. Figure 5 shows clusters identified within the measured data set projected onto a 2d plot. Black dots in the figure indicate outliers, while colored dots indicate unique clusters identified within the data set.

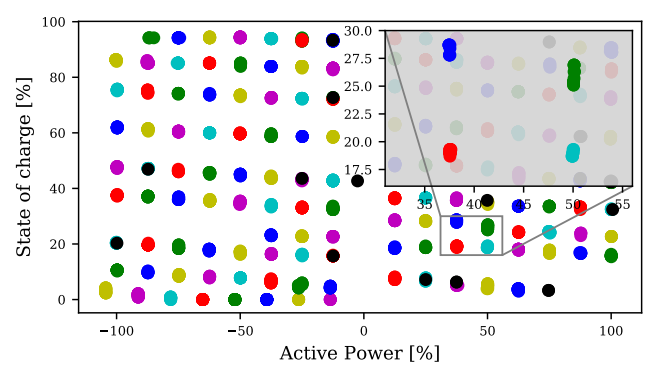

Fig. 5: Clusters form using DB-scan. Zoomed region shows that the algorithm does well identify unique clusters within noisy experimental data

Next, centroid of each cluster V_i is calculated using Equation 3, where N_i is the total number of points in the ith cluster and $i \in (1, 2, ..., k)$. Calculation of center of mass of each cluster allows mitigation of natural variance in experimental

data sets that originates from measurement error as well as variation in small variations in system state.

$$M(V_i) = \sum_{j=1}^{N_i} \frac{p_j}{N_i}; \tag{3}$$

where $p_1, p_2, ..., p_{N_i} \in V_i$ and $p_i = (P_m, S_m, L_m)$

In the final step a new convex hull $H(M(V))$ is formed around cluster centroids using Equation 1. Lower envelope of the hull is used as piece-wise linear model for estimation of system losses for a given value of active power and state of chart.

III. Model Validation

To demonstrate how the pre-cleaning via clustering procedure improves model accuracy, the linearized convex loss models both before and after cleaning is evaluated using two classes of analysis; residual analysis and global metrics. The residual analysis involes characterizing the first four moments of the distribution of the residuals of each model realtive to the measured data including the mean, variance, skewness and kurtosis. The global metrics include the root mean squared error (RMSE) and the coefficient of determination (R^2 score). RMSE quantifies the difference between measured and estimated values for total system losses and has been calculated using Equation 4.

$$RMSE(\overrightarrow{y}) = \sum_{i=1}^{n} \frac{\sqrt{y_i - \widehat{y}_i}}{n} \tag{4}$$

The R^2 score gauges how well the variance in the dependent variable is predicted from that of the independent variables. R^2 score is calculated using Equation 5.

$$R^2 score(\overrightarrow{y}) = \frac{\sum_{i=1}^{n} \sqrt{y_i - \widehat{y}_i}}{\sum_{i=1}^{n} \sqrt{y_i - \overline{y}_i}} \times 100\% \tag{5}$$

A. Residual analysis

The cumulative probability density of the residuals for each model are compared in Figure 6. Further, the distributions of the residuals for each model are compared in Figure 7 and the first four moments of the distributions are summarized in Table I. In this test case, both models systematically underestimate the total system losses as evidenced by the negative mean residual values, though the pre-cleaning via clustering procedure significantly reduces this mean bias error. The clustering procedure also reduces the variance of the residuals, but has little impact on the skewness of the distribution of residuals. The relatively large kurtosis associated with the distribution of residuals after the clustering procedure was applied indicates that the distribution is more "heavily-tailed", meaning that the distribution is more localized or "peaked" near the mean value.

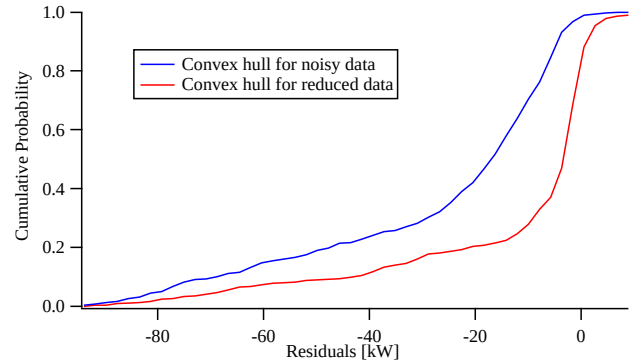

Fig. 6: Cumulative probability density plot of residuals of the convex hull models based on the noisy data and the reduced data set.

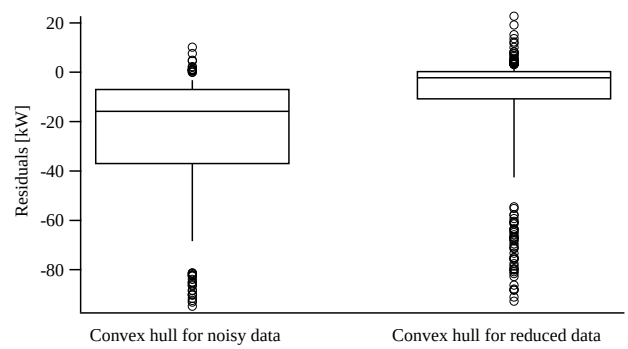

Fig. 7: Box and whisker plots of residuals of each model relative to the measured data.

TABLE I: The first four moments of the distribution of the residuals

Evaluation metric	Convex hull model for noisy data	Convex hull model for reduced data
mean	-25.279	-11.472
variance	604.914	466.690
skewness	-1.176	-1.989
kurtosis	0.247	3.160

B. Global accuracy metrics

The global accuracy metrics show that the clustering procedure results in a model which better predicts the total system losses of the flow battery. By reducing the clusters to representative centroids, the complexity of the developed piecewise linear model can be reduced. In this study, reduction of more than 27% in the number of hyperplanes that form the convex hull is seen (Table II). This reduces the complexity of the MILP implementations that makes use of the developed model.

One final observation that is worth a mention is that although measurement data set is predominantly convex, there are concave regions within the data. This is evident from Figure 1. At low values of active power a gap between the formed hull and the surface that is formed for the measured values is clearly visible. While the battery is operating within this region, model will systematically underestimated total system losses. This also explains the long tail in the cumulative probability plot in Figure 6.

TABLE II: Addition metrics use to to compare the developed models

Evaluation metric	**Convex hull model for noisy data**	**Convex hull model for reduced data**
R2 score [%]	90.1	95.3
Number of hyperplanes	190	138
Number of vertices	97	71

IV. CONCLUSIONS

In this paper, a novel method for developing piece-wise linear model using a convex hull is presented. The method makes use clustering techniques to remove outliers and mitigate measurement noise. Lower envelope of the convex hull formed using centroids of identified clusters has been used to define the piece-wise linear model. The algorithm is also capable of extrapolating outside defined region and estimating values for states that lie out the region enclosed by the hull. The proposed method has been compared to a convex hull model defined using the noisy data set directly. Comparison of results from the two models show that the proposed algorithm improves model accuracy significantly. In future work, efficacy of the proposed extrapolation method will be looked into in greater detail. Also, concave sub-regions with the measurement data set lead to systematic under estimation. the proposed method does not does not cater to this problem. Other methods such Delaunay triangulation that are capable of forming a non-convex hulls may useful to mitigate this issue. This will be also investigated in future work.

V. ACKNOWLEDGMENTS

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

REFERENCES

[1] J. A. Taylor, *Convex Optimization of Power Systems*. Cambridge University Press, feb 2015. [Online]. Available: https://www.cambridge.org/core/product/identifier/9781139924672/type/book

[2] E. Henry and J. Hofrichter, "[8] singular value decomposition: Application to analysis of experimental data," in *Numerical Computer Methods*, ser. Methods in Enzymology. Academic Press, 1992, vol. 210, pp. 129 – 192. [Online]. Available: http://www.sciencedirect.com/science/article/pii/007668799210010B

[3] S. Chen, S. Billings, and P. Grant, "Non-linear system identification using neural networks," *International journal of control*, vol. 51, no. 6, pp. 1191–1214, 1990.

[4] H. Ahmadi, J. R. Martí, and A. Moshref, "Piecewise Linear Approximation of Generators Cost Functions Using Max-Affine Functions," Tech. Rep. [Online]. Available: http://www.ece.ubc.ca/ hameda/papers/C4.pdf

[5] X. Gao, Z. Feng, Y. Wang, X. Huang, D. Huang, T. Chen, and X. Lian, "Piecewise Linear Approximation Based MILP Method for PVC Plant Planning Optimization," *Industrial and Engineering Chemistry Research*, vol. 57, no. 4, pp. 1233–1244, jan 2018. [Online]. Available: http://pubs.acs.org/doi/10.1021/acs.iecr.7b02130

[6] T. A. A. Victoire and A. E. Jeyakumar, "Hybrid pso–sqp for economic dispatch with valve-point effect," *Electric Power Systems Research*, vol. 71, no. 1, pp. 51–59, 2004.

[7] C. D'Ambrosio, A. Lodi, and S. Martello, "Piecewise linear approximation of functions of two variables in MILP models," *Operations Research Letters*, vol. 38, no. 1, pp. 39–46, jan 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167637709001072

[8] T. Simpkins, D. Cutler, K. Anderson, D. Olis, E. Elgqvist, M. Callahan, and A. Walker, "REopt: A Platform for Energy System Integration and Optimization," in *Volume 2: Economic, Environmental, and Policy Aspects of Alternate Energy; Fuels and Infrastructure, Biofuels and Energy Storage; High Performance Buildings; Solar Buildings, Including Solar Climate Control/Heating/Cooling; Sustainable Cities and Communit*. ASME, jun 2014, p. V002T03A006. [Online]. Available: http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?doi=10.1115/ES2014-6570

[9] J. M. D. Sejdić, A. Kurosu, J. L. Coyle, and Ervin, "A comparative analysis of DBSCAN, K-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals," *Computers in Biology and Medicine*, vol. 59, pp. 10 – 18, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482515000244

[10] X. and others Ester, Martin and Kriegel, Hans-Peter and Sander, J{\"o}rg and Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." *Kdd*, vol. 96, no. 34, pp. 226—-231, 1996.

[11] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.